The Effect of Affect: Modeling the Impact of Emotional State on the Behavior of Interactive Virtual Humans

Stacy Marsella,[1] Jonathan Gratch,[2] and Jeff Rickel[1]

[1] USC Information Sciences Institute
4676 Admiralty Way, Suite 1001, Marina del Rey, CA, 90292
marsella@isi.edu, rickel@isi.edu

[2] USC Institute for Creative Technologies
13274 Fiji Way, Suite 600, Marina del Rey, CA, 90292
gratch@ict.usc.edu

ABSTRACT

A person's behavior provides significant information about their emotional state, attitudes, and attention. Our goal is to create virtual humans that convey such information to people while interacting with them in virtual worlds. The virtual humans must respond dynamically to the events surrounding them, which are fundamentally influenced by users' actions, while providing an illusion of human-like behavior. A user must be able to interpret the dynamic cognitive and emotional state of the virtual humans using the same nonverbal cues that people use to understand one another. Towards these goals, we are integrating and extending components from three prior systems: a virtual human architecture with a range of cognitive and motor capabilities, a model of emotional appraisal, and a model of the impact of emotional state on physical behavior. We describe the key research issues, our approach, and an initial implementation in an Army peacekeeping scenario.

1. INTRODUCTION

A person's emotional state influences them in many ways. It impacts their decision making, actions, memory, attention, voluntary muscles, etc., all of which may subsequently impact their emotional state (e.g., see [4]). This pervasive impact is reflected in the fact that a person will exhibit a wide variety of nonverbal behaviors consistent with their emotional state, behaviors that can serve a variety of functions both for the person exhibiting them as well as for people observing them. For example, shaking a fist at someone plays an intended role in communicating information. On the other hand, behaviors such as rubbing one's thigh, averting gaze, or a facial expression of fear may have no explicitly intended role in communication. Nevertheless, these actions may suggest considerable information about a person's emotional arousal, their attitudes, and their focus of attention.

Our goal is to create virtual humans that convey these types of information to humans while interacting with them in virtual worlds. We are interested in virtual worlds that offer human users an engaging scenario through which they will gain valuable experience. For example, a young Army lieutenant could be trained for a peacekeeping mission by putting him in virtual Bosnia and presenting him with the sorts of situations and dilemmas he is likely to face. In such scenarios, virtual humans can play a variety of roles, such as an experienced sergeant serving as a mentor, soldiers serving as his teammates, and the local populace. Unless the lieutenant is truly drawn into the scenario, his actions are unlikely to reflect the decisions he will make under stress in real life. The effectiveness of the training depends on our success in creating engaging, believable characters that convey a rich inner dynamics that unfolds in response to the scenario.

Thus, our design of the virtual humans must satisfy three requirements. First, they must be believable; that is, they must provide a sufficient illusion of human-like behavior that the human user will be drawn into the scenario. Second, they must be responsive; that is, they must respond to the events surrounding them, which will be fundamentally influenced by the user's actions. Finally, they must be interpretable; the user must be able to interpret their response to situations, including their dynamic cognitive and emotional state, using the same nonverbal cues that people use to understand one another. Thus, our virtual humans cannot simply create an illusion of life through cleverly designed randomness in their behavior; their inner behavior must respond appropriately to a dynamically unfolding scenario, and their outward behavior must convey that inner behavior accurately and clearly.

This paper describes our progress towards a model of the outward manifestations of an agent's emotional state. Our work integrates three previously implemented systems. The first, Steve [18, 20, 19], provides an architecture for virtual humans in 3D simulated worlds that can interact with human users as mentors and teammates. Although Steve did not include any emotions, his broad capabilities provide a foundation for the virtual humans towards which we are working. The second, Émile [8], focuses on emotional appraisal: how emotions arise from the relationship between environmental events and an agent's plans and goals. The third, IPD [11], contributes a complementary model of emotional appraisal as well as a model of the impact of emotional state on physical behavior. The integration of these three systems provides an initial model of virtual humans for experiential learning in engaging virtual worlds. Our work is part of a larger effort to add a variety of new capabilities to Steve, including more sophisticated support for spoken dialog and a more human-like model of perception [17].

Report Documentation Page

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **2001**	2. REPORT TYPE	3. DATES COVERED **00-00-2001 to 00-00-2001**

4. TITLE AND SUBTITLE **The Effect of Affect: Modeling the Impact of Emotional State on the Behavior of Interactive Virtual Humans**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California,Information Sciences Institute ,4676 Admiralty Way,Marina del Rey,CA,90292**	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**5**	

Figure 1: An interactive peacekeeping scenario featuring (left to right) a sergeant, a mother, and a medic

2. EXAMPLE SCENARIO

To illustrate our vision for virtual humans that can interact with people in virtual worlds, we have implemented an Army peacekeeping scenario, which has been viewed by several hundred people and was favorably viewed by many domain experts [21]. As the simulation begins, a human user, playing the role of a U.S. Army lieutenant, finds himself in the passenger seat of a simulated vehicle speeding towards a Bosnian village to help a platoon in trouble. Suddenly, he rounds a corner to find that one of his platoon's vehicles has crashed into a civilian vehicle, injuring a local boy (Figure 1). The boy's mother and an Army medic are hunched over him, and a sergeant approaches the lieutenant to brief him on the situation. Urgent radio calls from the platoon downtown, as well as occasional explosions and weapons fire from that direction, suggest that the lieutenant send his troops to help them. Emotional pleas from the boy's mother, as well as a grim assessment by the medic that the boy needs a medevac immediately, suggest that the lieutenant instead use his troops to secure a landing zone for the medevac helicopter. The lieutenant carries on a dialog with the sergeant and medic to assess the situation, issue orders (which are carried out by the sergeant through four squads of soldiers), and ask for suggestions. His decisions influence the way the situation unfolds, culminating in a glowing news story praising his actions or a scathing news story exposing the flaws in his decisions and describing their sad consequences.

While the current implementation of the scenario offers little flexibility to the lieutenant, it provides a rich test bed for our research. Currently, the mother, medic and sergeant are implemented as Steve agents. All other virtual humans (a crowd of locals and four squads of soldiers) are scripted characters. We will illustrate elements of our design using the mother, since her emotions are both the most dramatic in the scenario and the most crucial for influencing the lieutenant's decisions and emotional state. While our current implementation of the mother includes a preliminary integration of Steve, Émile, and IPD, the focus of this paper is on a more general integration that goes beyond that implementation.

3. RELATED WORK

Several researchers have considered computational models of emotion to drive the behavior of synthetic characters. These approaches can be roughly characterized as being either communication driven or appraisal driven. Communication driven means that one selects a communication act and then chooses emotional expression based on some desired impact it will have on the user. For example, Cosmo [10] makes a one-to-one mapping between speech acts and emotional behaviors that reinforce the act (congratulatory acts trigger an admiration emotive intent that is conveyed with applause). Ball and Breese [2] intend to convey a sense of empathy by expressing behaviors that mirror the assessed emotional state of the user. Poggi and Pelachaud [16] use facial expressions to convey the performative of a communication act, showing "potential anger" to communicate that the agent will be angry if a request is not fulfilled.

In contrast, appraisal theories focus on the apparent evaluative function emotions play in human reasoning. Appraisal theories view emotion as arising from some assessment of an agent's internal state vis-à-vis its external environment (e.g., is this event contrary to my desire?). Such appraisals can be used to guide decision-making and behavior selection. For example, Beaudoin [3] uses them to set goal priorities and guide meta-planning. They can also serve as a basis for communicating information about the agent's assessment, though not in the intentional way viewed by communicative models. Appraisal methods are useful for giving coherence to an agent's emotional dynamics that can be lacking in purely communicative models. This coherence is essential for conveying a sense of believability and realism [13].

These approaches are complementary, though few approaches have considered how to combine them into a coherent approach. An exception is IPD [11], which represents both appraised emotional state and communicative intent and selectively expresses one or the other based on a simple threshold model. In the discussion that follows we will focus on the problem of emotional appraisal.

4. MODELING EMOTIONS
4.1 Emotional State

Our primary design goal is to create general mechanisms that are not tied to the details of any specific domain. Thus, emotional appraisals do not refer to features of the environment directly, but reference abstract features of the agent's information processing. Appraisals and re-appraisals are triggered reflexively by changes

in an agent's mental representations, which are in turn changed by general reasoning mechanisms. External events can indirectly impact these internal representations. For example, the agent may perceive an event (e.g., soldiers are leaving the accident scene), and then a general reasoning mechanism (e.g. Steve's task reasoning) infers the consequences of this event (e.g., the departure of troops violates the precondition of them treating my child). Alternatively, these representations may change as the result of internal decision-making (e.g., forming an intention to violate a social obligation). Appraisal frames key off of features of these representations (e.g., the negative impact of a precondition violation on goal probability or the violation of the social norm of meeting your obligations). The dynamics of behavior and emotional expression is thereby driven by the mechanisms that update an agent's mental representations.

The Émile system [8] illustrates this approach by using a domain-independent planning approach to represent mental state and make inferences, but Émile is therefore largely restricted to appraisals related to an agent's goals and changes to the probability of goal attainment. IPD [11] explores a much richer notion of appraisal including a variety of emotional coping strategies and a richer notion of social relationships and ego identity, but is implemented in terms of a less general inference mechanism. In our joint work, we begin to bridge the gap between by buttressing Émile-style appraisals with a broader repertoire of inference strategies. Initially, we are augmenting plan-based appraisals with those related to dialog, incorporating methods for representing and updating the discourse state and social obligations that arise from discourse [15]. These structures can then serve as the basis for a more uniform representation of social commitments and obligations beyond those strictly related to dialog, and that interact with plan-based reasoning (e.g. one is obligated to redress wrongs). This will allow the system to model some of the rich dynamics of emotions relating to discourse that IPD supports, but do so in a more structured way.

Émile contains a set of recognition rules that scan an agent's internal representations and generate an appraisal frame whenever certain features are recognized. For example, whenever the agent adopts a new goal (or is informed of a goal of some other agent), frames are created to track the status of that goal. Each appraisal frame describes the appraised situation in terms of a number of specific features proposed by the OCC model [14]. These include the point of view from which the appraisal is formed, the desirability of the situation, whether the situation has come to pass or is only a possibility, whether the situation merits praise or blame (we do not model the appealingness of domain objects, an important factor in OCC). For example, if someone threatens my goal, from my perspective this is an undesirable possibility that merits blame. Based on the setting of the various features, one or more "emotion instances" are generated according to OCC. Émile incorporates a decision-theoretic method for assigning intensities to these instances. Intensity values decay over time but are "re-energized" whenever the underlying representational structures are manipulated.

Individual emotion instances are aggregated into "buckets" corresponding to emotions of the same type. Thus, threats to multiple goals will be aggregated into an overall level of fear. In this sense, Émile contains a two-level representation of emotional state. Emotion instances are the level at which some semantic connection is still retained between the emotion and the mental structures underlying the appraisal. Thus, the agent has the capability of answering questions about its emotion state (e.g., "Why are you so angry?"). The aggregate buckets roughly correspond to the over-

all assessment of the agent's emotional state and are used to drive the physical focus modes discussed below. This model does yet support some of the subtle dynamics displayed by IPD which can convey an overall tone yet make agile transitions between conveyed emotional state. We are exploring some different strategies for achieving this effect, including aggregating subsets of appraisal frames on demand depending of the current discourse state, or by adding another aggregate layer that incorporates a different decay rate into appraisal intensities.

4.2 The Effect of Emotions on Behavior

The agents we design incorporate a wide range of outward behaviors in order to interact believably with the environment as well as other agents and humans. Their bodies have fully articulated limbs, facial expressions, and sensory apparatus. They can move in the environment, manipulate objects and direct their gaze in appropriate ways. They are capable of rich, multi-modal communication that incorporates both verbal behaviors as well as nonverbal behaviors. In addition, they have facial expressions, body postures and the ability to perform various kinds of gestures. The key challenge for the agent design is to manage this flexibility in the agent's physical presence in a way that conveys consistent emotional state and individual differences. To address these challenges, we rely on a wide range of work in human emotion, social behavior, clinical psychology, animation and the performance arts.

For example, psychological research on emotion reveals its pervasive impact on physical behavior such as facial expressions, gaze and gestures [1, 5, 6]. These behaviors suggest considerable information about emotional arousal, attitudes and attention. For example, depressed individuals may avert their gaze and direct their gestures inward towards their bodies in self-touching behaviors.

Note that such movements also serve to mediate the information available to the individual. For example, if a depressed individual's head is lowered, this also regulates the information available to the individual. Orienting on an object of fear or anger brings the object to the focus of perceptual mechanisms, which may have indirect influences on cognition and cognitive appraisal by influencing the content of working memory. Even a soothing behavior like rubbing an arm may serve to manage what a person attends to [7].

These findings provide a wealth of data to inform agent design but often leave open many details as to how alternative behaviors are mediated. The agent technology we use allows one to create rich physical bodies for intelligent characters with many degrees of physical movement. This forces one to directly confront the problem of emotional consistency. For example, an "emotionally depressed" agent might avert gaze, be inattentive, perhaps hugging themselves. However, if in subsequent dialog the agent used strong communicative gestures such as beats [12], then the behavior might not "read" correctly. This issue is well known among animators. For example, Frank Thomas and Ollie Johnston, two of Disney's "grand old men," observe that "the expression must be captured throughout the whole body as well as in the face" and "any expression will be weakened greatly if it is limited only to the face, and it can be completely nullified if the body or shoulder attitude is in any way contradictory" [22].

Implicit in these various concerns is that the agent has what amounts to a resource allocation problem. The agent has limited physical assets, e.g., two hands, one body, etc. At any point in time, the agent must allocate these assets according to a variety of de-

mands, such as performing a task, communicating, or emotionally soothing themselves. For instance, the agent's dialog may be suggestive of a specific gesture for the agent's arms and hands while the emotional state is suggestive of another. The agent must mediate between these alternative demands in a fashion consistent with their goals and their emotional state. From the human participant's perspective, we expect this consistency in the agent's behavior to support believability and interpretability.

To address this problem we rely on the Physical Focus model [11], a computational technique inspired by work on nonverbal behavior in clinical settings [7] and Lazarus's [9] delineation of emotion-directed versus problem-directed coping strategies. The Physical Focus model bases an agent's physical behavior in terms of what the character attends to and how they relate to themselves and the world around them, specifically whether they are focusing on themselves and thereby withdrawing from the world or whether they are focusing on the world, engaging it. In other words, the model bases physical behavior on how the agent is choosing to cope with the world.

The model organizes possible behaviors around a set of modes. Behaviors can be initiated via requests from the planner/executor or started spontaneously when the body is not otherwise engaged. At any point in time, the agent will be in a unique mode based on the current emotional state.

In the current work, we model three modes of physical focus: Body Focus, Transitional Focus and Communicative Focus (as opposed to the five modes discussed in [11] and identified in Freedman's work). The focus mode impacts behavior along several key dimensions, including which behaviors are available, how behaviors are performed, how attentive to external events the agent is, and which actions the agent selects and how it speaks.

The Body Focus mode models a self-focused attention, away from the world, and is also designed to reveal considerable depression or guilt. The prevalent type of gestures available to the agent in this mode are self-touching gestures, or adaptors [5], which involve hand-to-body gestures that appear soothing (e.g., rhythmic stroking of forearm) or self-punitive (e.g., squeezing or scratching of forearm). Conversely, communicative gestures such as deictic or beat gestures [12] are muted in this mode both in terms of their numbers and dynamic extent. Indeed, overall gesturing in this mode is muted in its dynamics. Also, the agent is less attentive to the environment and in particular exhibits considerable gaze aversion and downward looking gaze. The agent's verbal behavior is inhibited and marked by pauses. In terms of action selection, the agent has a reduced preference for communicative acts.

Transitional Focus indicates a less divided attention, less depression or guilt, a burgeoning willingness to take part in the conversation, milder conflicts with the problem solving and a closer relation to the listener. This mode's gesturing is marked by hand-to-hand gestures (such as rubbing hands or hand fidgetiness) and hand-to-object gestures, such as playing with a pen. There are more communicative gestures in this mode but they are still muted or stilted.

Finally, Communicative Focus is designed to indicate a full willingness to engage the world in dialog and action. In this mode, the agent's full range of communicative gestures are available. On the other hand, adaptors are muted and occur only when the agent is listening or otherwise not occupied. In contrast, the communicative gestures are more exaggerated, both in terms of numbers

and physical dynamics. The agent is also more attentive to and responsive to events in the environment.

Transition between modes is currently based on emotional state derived from the appraisal model. High levels of sadness, decreased hope, or guilt, in absolute terms and relative to other emotional buckets, induces transitions away from Communicative Focus and towards Body Focus. Increased hope or anger induces transitions towards Communicative Focus. The transitions are designed to be "sticky" so that the agent does not readily pop into and then out of a mode. Transitional Focus, true to its name, lies between the other two modes. However, extreme emotional state can bypass it.

Grouping behaviors into modes attempts to (a) mediate competing demands on an agent's physical resources and (b) coalesce behaviors into groups that provide a consistent interpretation about how the agent is relating to its environment. It is designed to do this in a fashion consistent with emotional state, with the intent that it be general across agents. However, realism also requires that specific behaviors within each mode incorporate individual differences, as in human behavior. For example, we would not expect a mother's repertoire of gestures to be identical to that of an army sergeant. Indeed, the current implementation is a step en route to a fuller rendering of a coping model: a model of how the agent chooses to cope with emotional stress that includes both emotional appraisal as well as individual differences (personality).

5. CONCLUSION

The challenge of this research is not an easy one. We seek to create virtual humans whose behavior is a believable rendition of human-like behavior, is responsive to dynamically unfolding scenarios and is interpretable using the same cues that people use to understand each other. A key factor for success will be a faithful rendering of the cause of emotion and its outward manifestation. To address that challenge, we have begun to integrate components from three previous systems: Steve's virtual human architecture, Émile's emotional appraisal and IPD's complementary model of emotional appraisal and its impact on physical behavior.

An initial integration has already been realized and implemented in the mother character of our Army peacekeeping scenario. However, our integration efforts are ongoing. Our current efforts are on generalizing that integration and incorporating more of the IPD social-based appraisals into the general inferencing mechanisms in Steve and Émile.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] M. Argyle and M. Cook. *Gaze and Mutual Gaze*. Cambridge University Press, Cambridge, 1976.

[2] G. Ball and J. Breese. Emotion and personality in a conversational agent. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000.

[3] L. Beaudoin. *Goal Processing in Autonomous Agents*. PhD thesis, University of Birmingham, 1995. CSRP-95-2.

[4] L. Berkowitz. *Causes and Consequences of Feelings*. Cambridge University Press, 2000.

[5] P. Ekman and W. V. Friesen. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1:49–97, 1969.

[6] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *Personality and Social Psychology*, 17(2), 1971.

[7] N. Freedman. The analysis of movement behavior during the clinical interview. In A. Siegman and B. Pope, editors, *Studies in Dyadic Communication*, pages 177–210. Pergamon Press, New York, 1972.

[8] J. Gratch. Émile: Marshalling passions in training and education. In *Proceedings of the Fourth International Conference on Autonomous Agents*, pages 325–332, New York, 2000. ACM Press.

[9] R. Lazarus. *Emotion and Adaptation*. Oxford Press, 1991.

[10] J. C. Lester, J. L. Voerman, S. G. Towns, and C. B. Callaway. Deictic believability: Coordinating gesture, locomotion, and speech in lifelike pedagogical agents. *Applied Artificial Intelligence*, 13:383–414, 1999.

[11] S. C. Marsella, W. L. Johnson, and C. LaBore. Interactive pedagogical drama. In *Proceedings of the Fourth International Conference on Autonomous Agents*, pages 301–308, New York, 2000. ACM Press.

[12] D. McNeill. *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press, 1992.

[13] W. S. Neal Reilly. *Believable Social and Emotional Agents*. PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1996. Technical Report CMU-CS-96-138.

[14] A. Ortony, G. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.

[15] M. Poesio and D. Traum. Towards an axiomatization of dialogue acts. In J. Hulstijn and A. Nijhol, editors, *Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues (13th Twente Workshop on Language Technology)*, pages 207–222, Enschede, 1998.

[16] I. Poggi and C. Pelachaud. Emotional meaning and expression in performative faces. In *International Workshop on Affect in Interactions: Towards a New Generation of Interfaces*, Siena, Italy, 1999.

[17] J. Rickel, J. Gratch, R. Hill, S. Marsella, and W. Swartout. Steve goes to Bosnia: Towards a new generation of virtual humans for interactive experiences. In *AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment*, March 2001.

[18] J. Rickel and W. L. Johnson. Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence*, 13:343–382, 1999.

[19] J. Rickel and W. L. Johnson. Virtual humans for team training in virtual reality. In *Proceedings of the Ninth International Conference on Artificial Intelligence in Education*, pages 578–585. IOS Press, 1999.

[20] J. Rickel and W. L. Johnson. Task-oriented collaboration with embodied agents in virtual worlds. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000.

[21] W. Swartout, R. Hill, J. Gratch, W. Johnson, C. Kyriakakis, C. LaBore, R. Lindheim, S. Marsella, D. Miraglia, B. Moore, J. Morie, J. Rickel, M. Thiébaux, L. Tuch, R. Whitney, and J. Douglas. Toward the holodeck: Integrating graphics, sound, character and story. In *Proceedings of the Fifth International Conference on Autonomous Agents*, 2001.

[22] F. Thomas and O. Johnston. *The Illusion of Life: Disney Animation*. Walt Disney Productions, New York, 1981.